

Cross-species Comparison

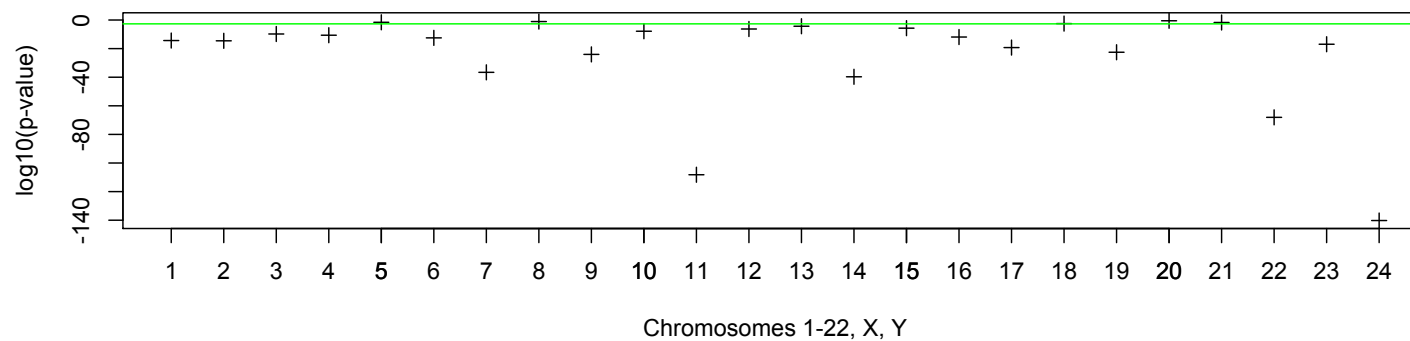
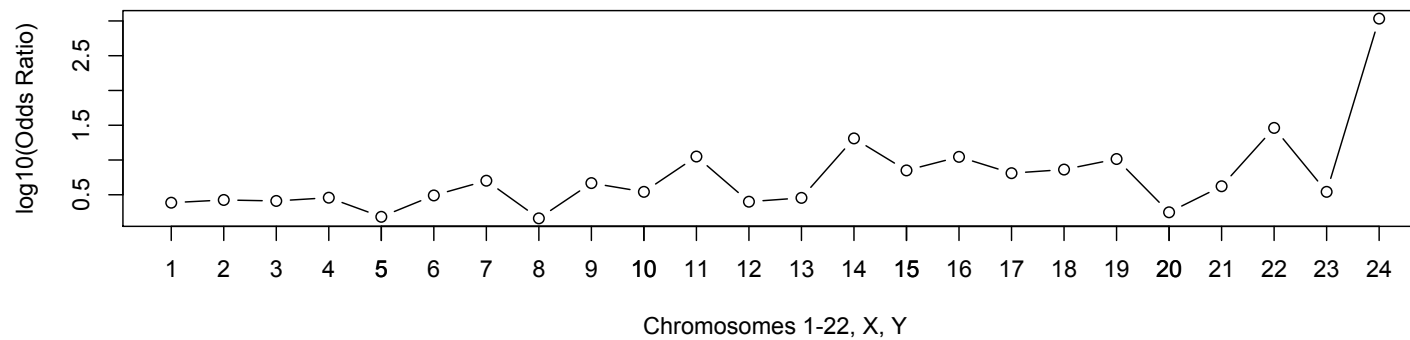
Pseudogene's Localization Relative to their Parent Genes

Yan Zhang

7/10/2013

Is co-residence random? Human: No.

Fisher's Exact Test: Co-residence is Random or Not for Each Chromosome



Worm, Fly: No, too.

Co-localization

- Co-residence: pseudogene-parent on the same protein
- Co-localization, distance threshold:
 - $D = 25\text{K}, 50\text{K}, 75\text{K}$ and 100K (bp) [1].
 - 25K is $1/8$ of average distance between genes [2].

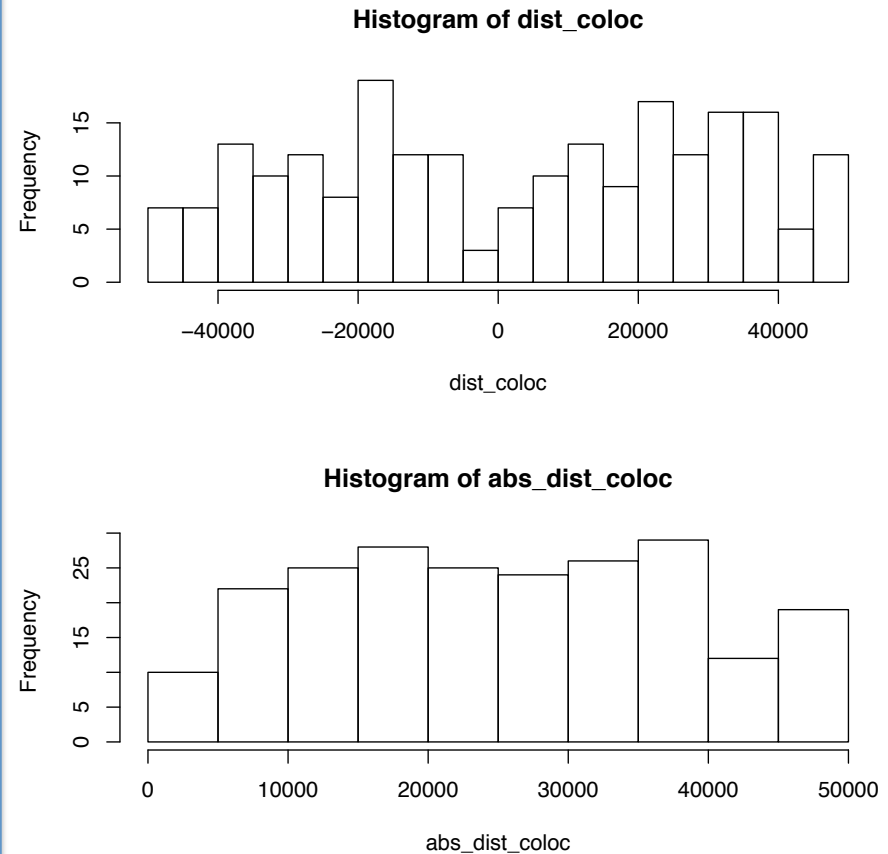
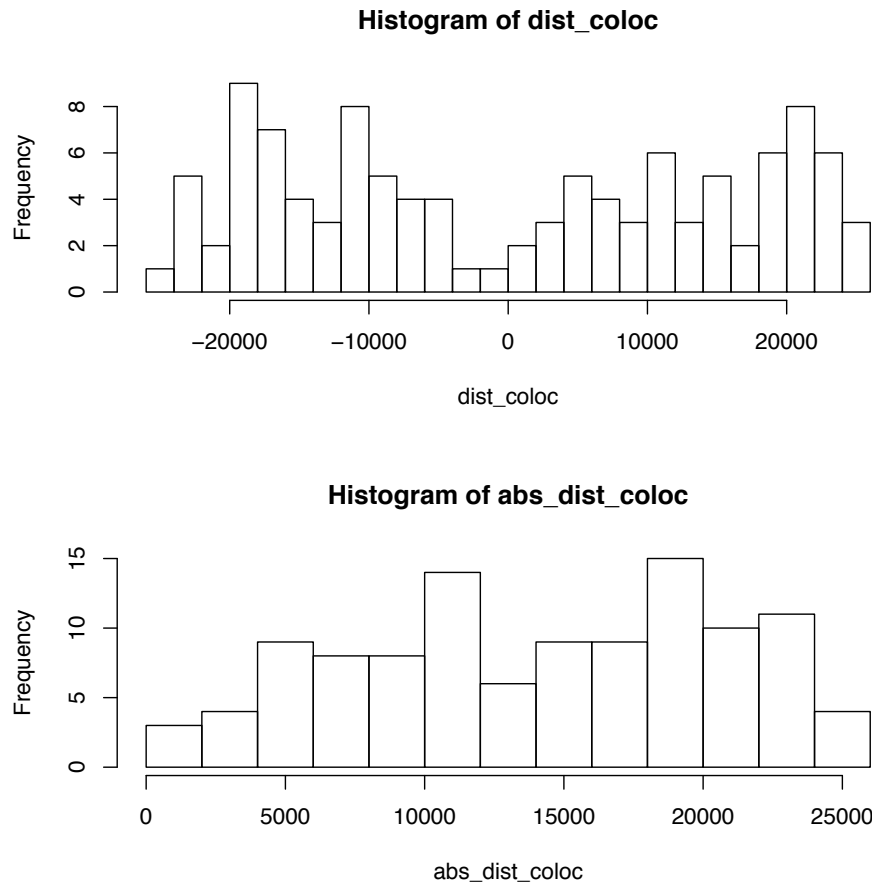
Cross-species comparison

- Human
- Worm
- Fly

Distribution of distance

Human Distribution of Distance (D=25K, 50K)

As seen on before:

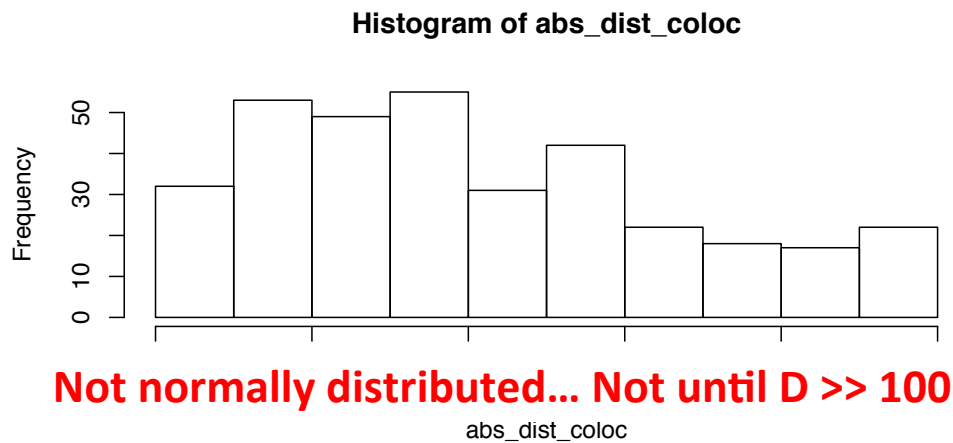
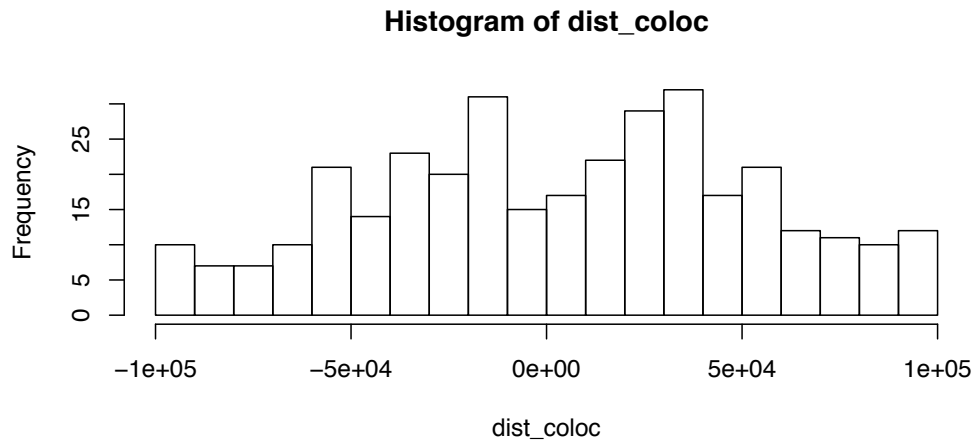


Not normally distributed... Not until $D \gg 100K$.

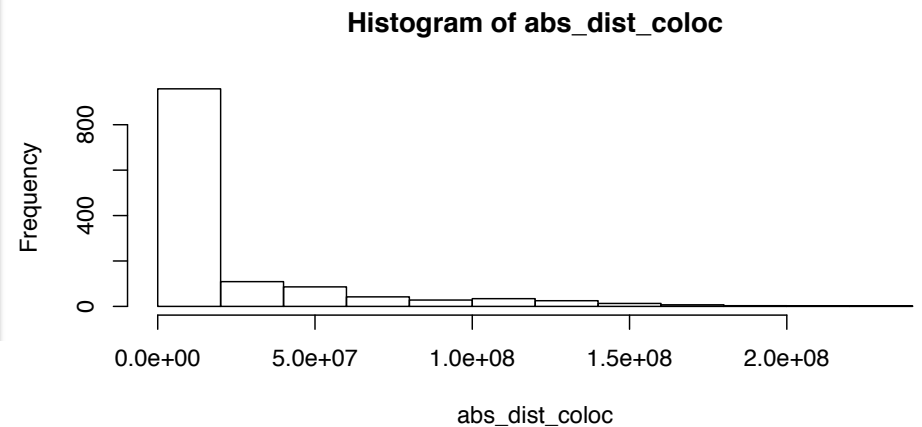
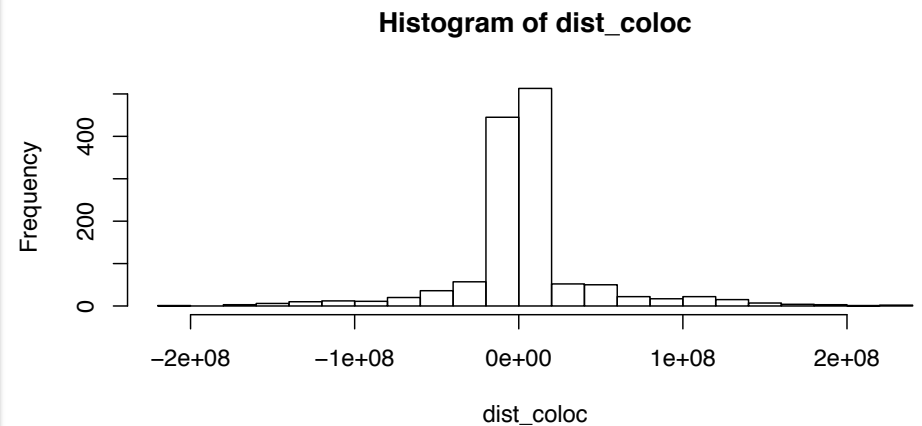
Human

Distribution of Distance (left: D=100K, right: Co-residence)

As seen before:

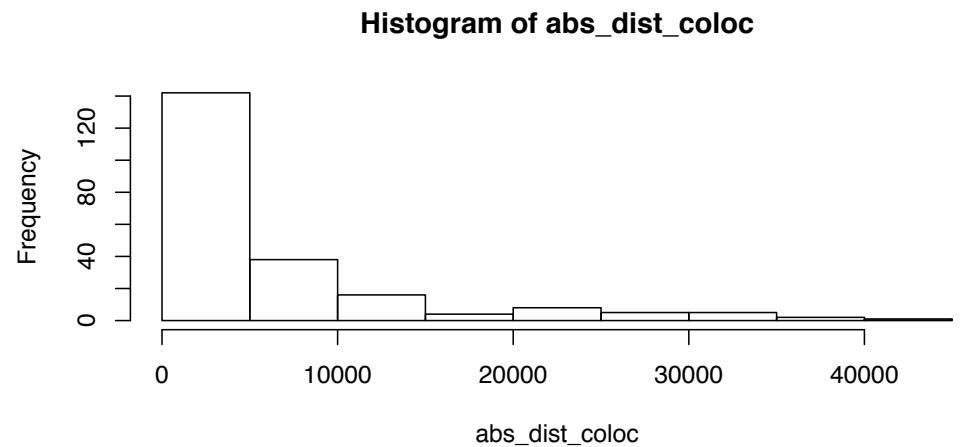
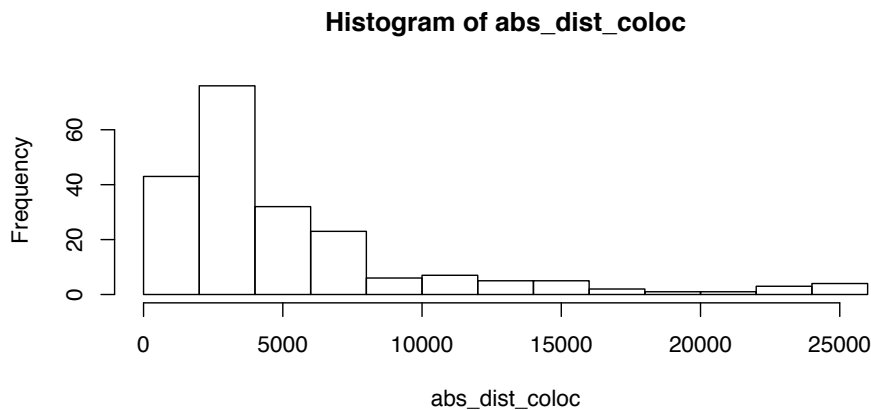
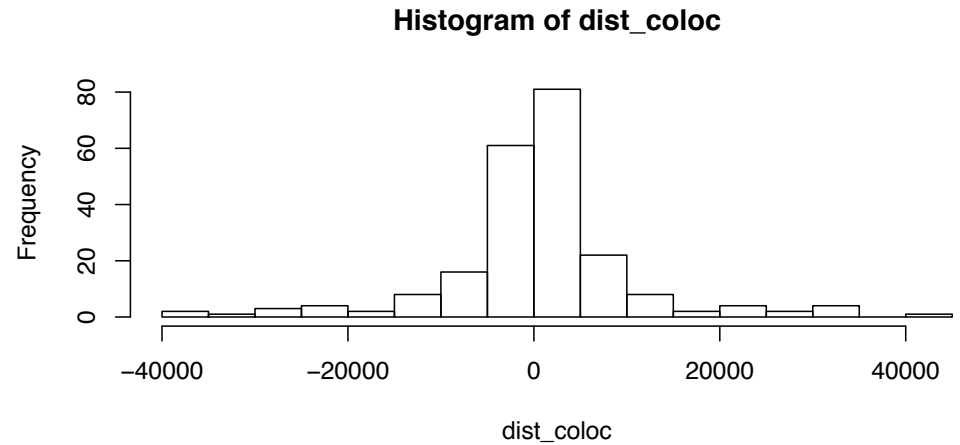
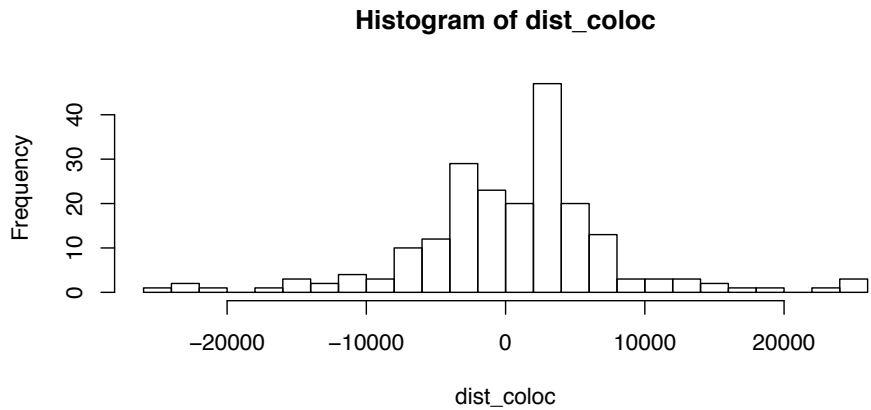


Not normally distributed... Not until $D \gg 100K$.



Worm

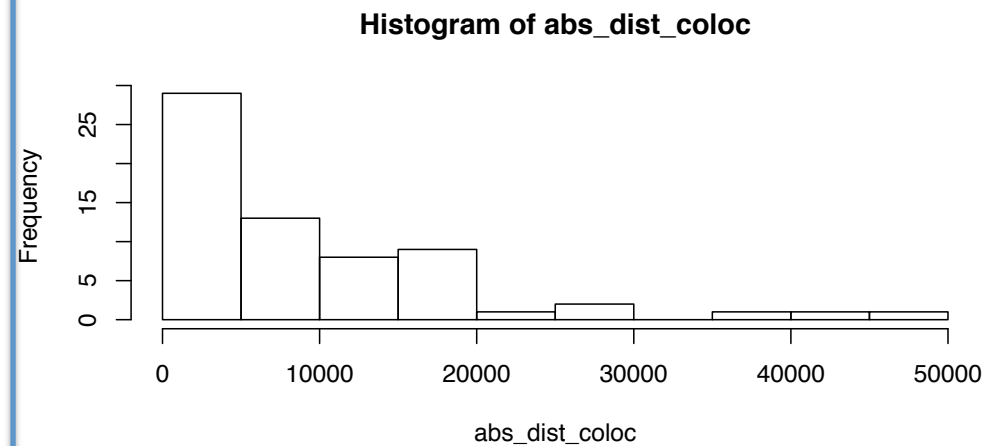
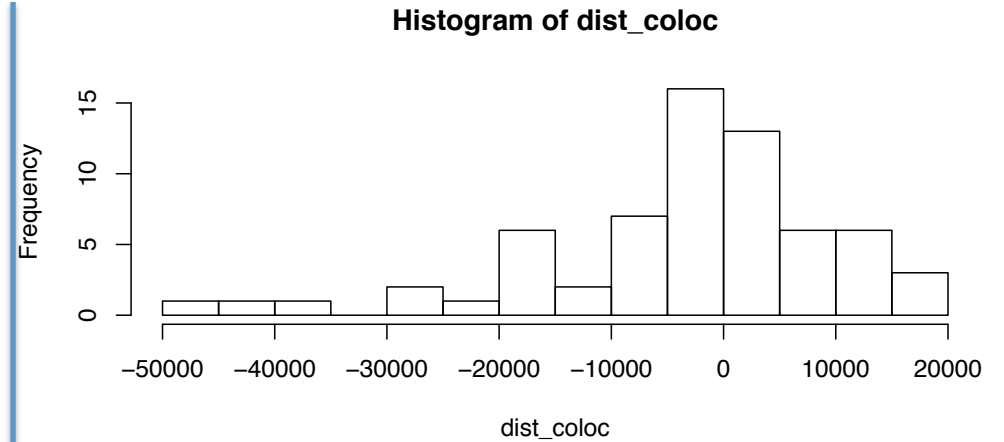
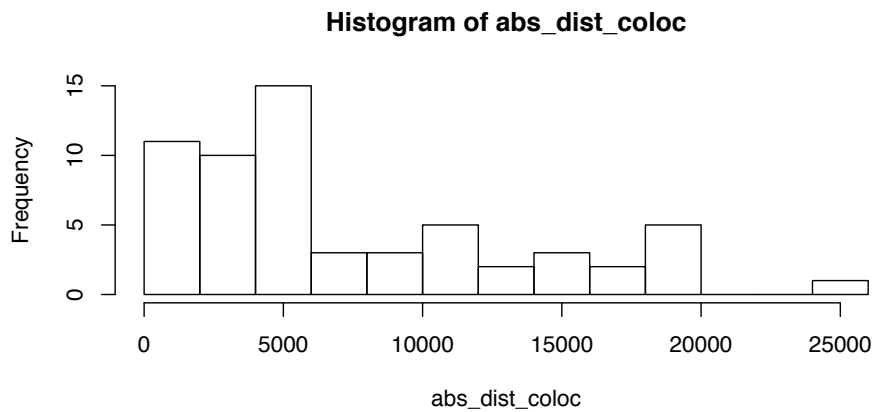
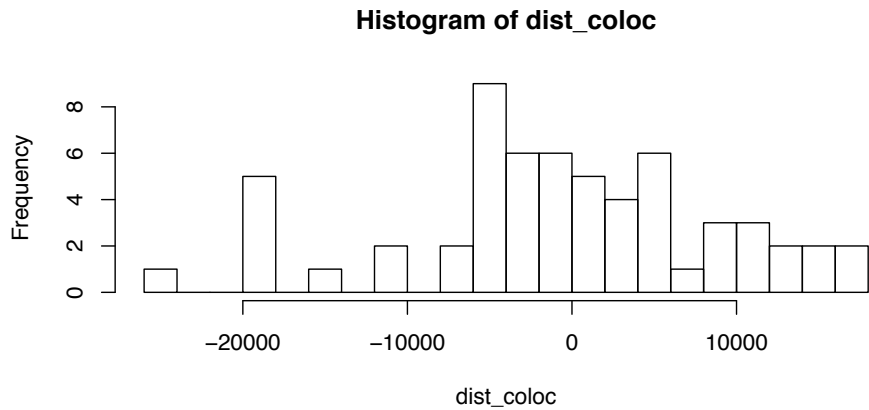
Distribution of Distance (D=25K, 50K)



More normally distributed than Human. But still shows heavy tailed pattern when D increases

Fly

Distribution of Distance (D=25K, 50K)



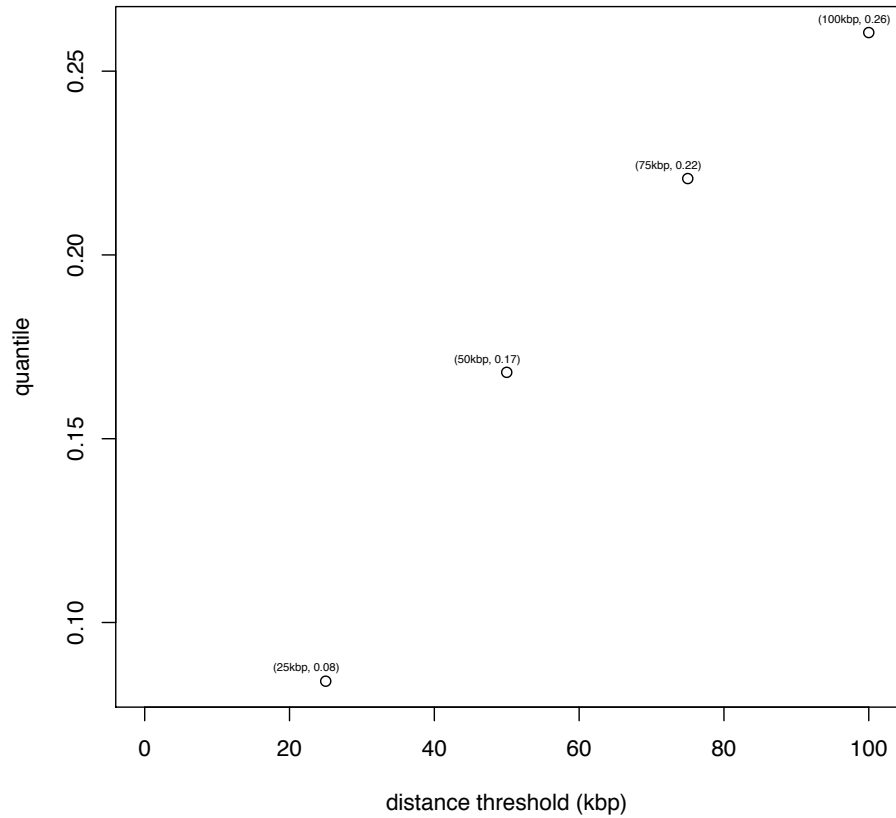
More normally distributed than Human. (QQplots available)

Quantiles of Ds

- X axis: D=25K, 50K, 75K, 100K
- Y axis: quantiles (of the distribution of distance as long as pairs on the same chromosomes)

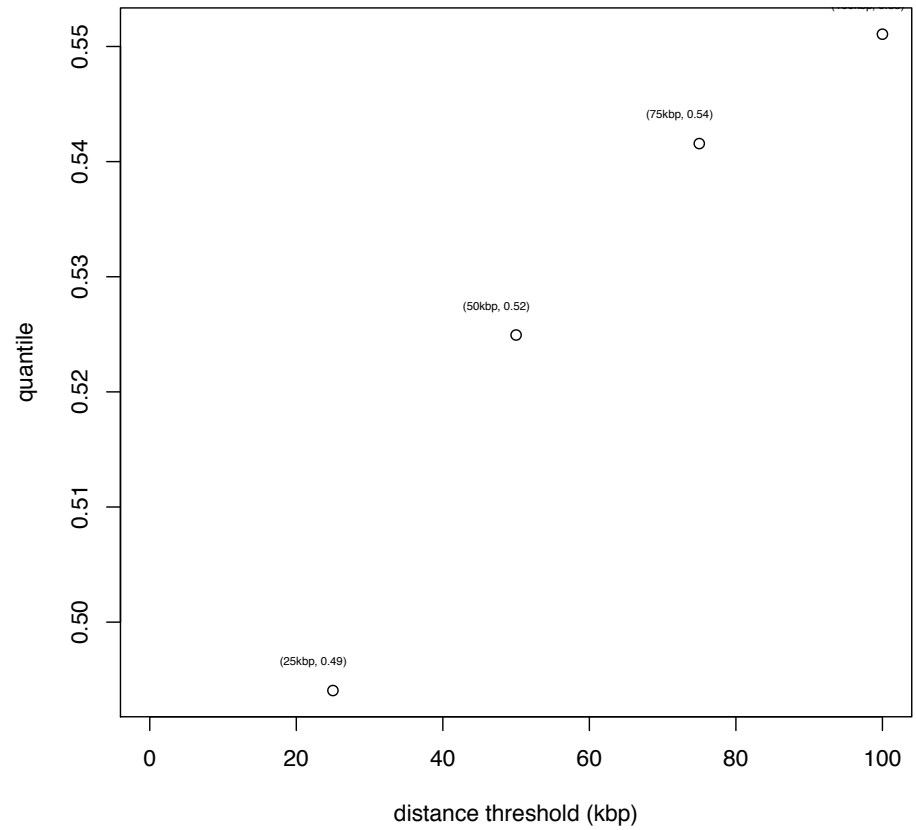
Quantiles of Ds

Quantiles of pseudogene–parent distance (on the same chromosome)



Human

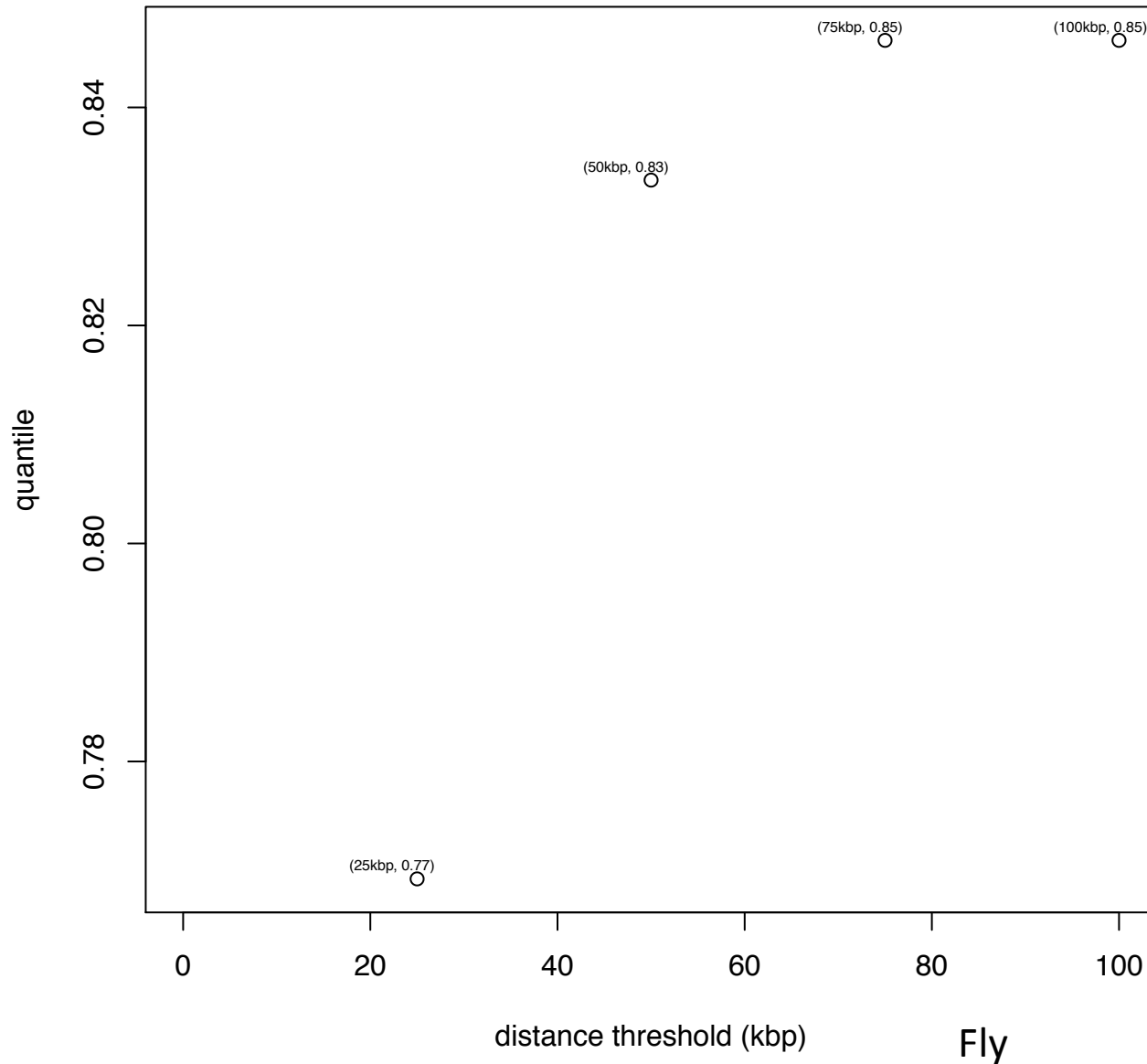
Quantiles of pseudogene–parent distance (on the same chromosome)



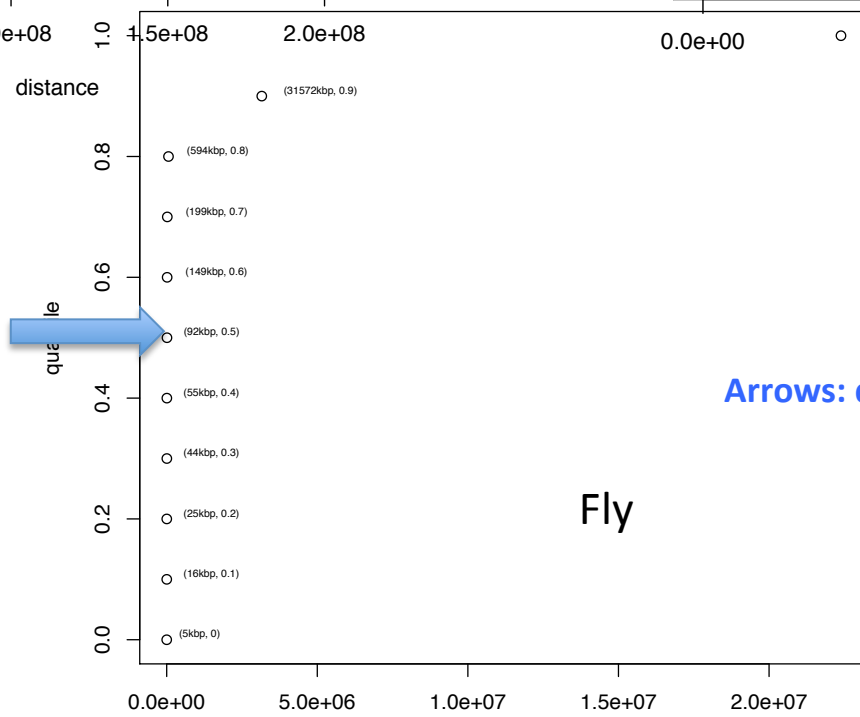
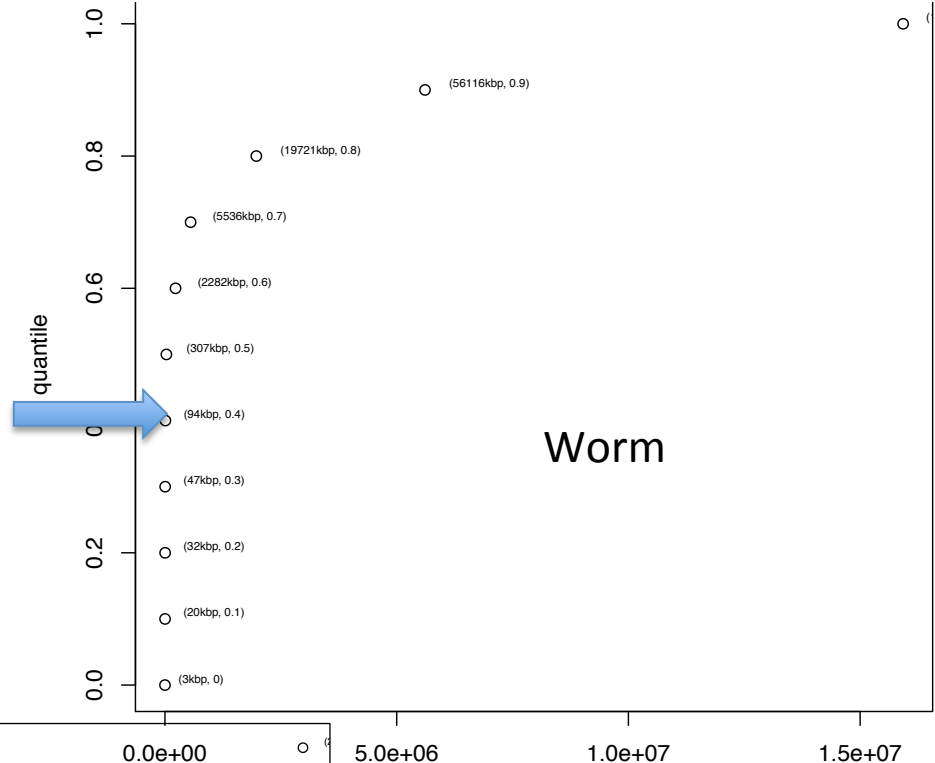
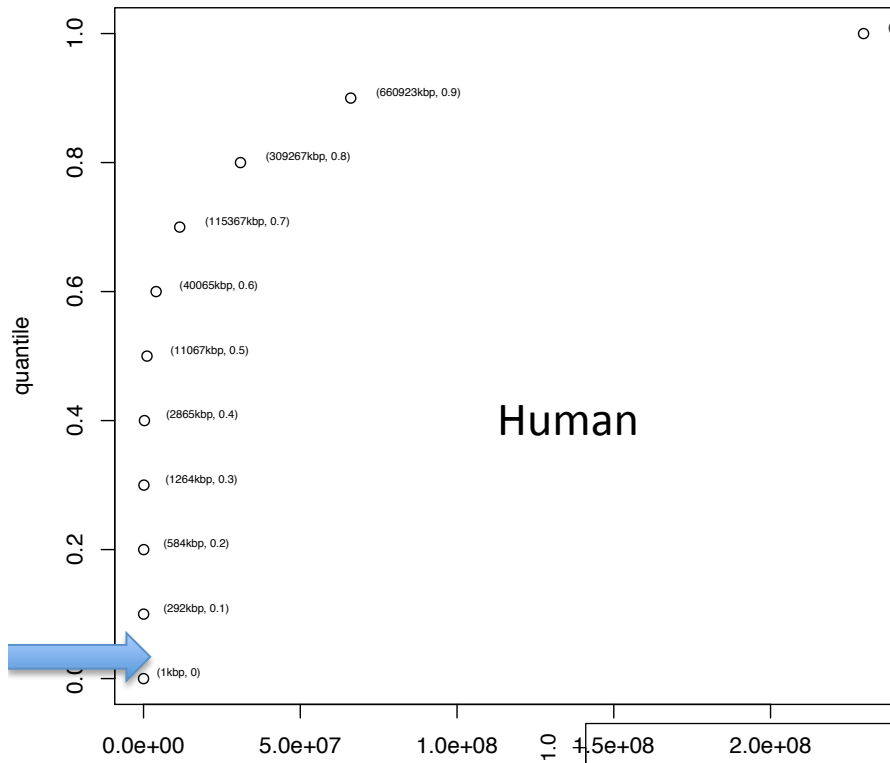
Worm

Quantiles of Ds

Quantiles of pseudogene–parent distance (on the same chromosome)



Distribution of Distance



Arrows: quantile corresponding to ~100K

Fitted Models

Human - Fitted Models (D=25K)

logit(Co-Loc25K) ~ Type * Transcribed + ParentChrom

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.26479    0.33037  -6.855 7.12e-12 ***
TypePSSD      -3.05381    0.26545 -11.504 < 2e-16 ***
TranscribedTRANS_YES  0.26209    0.30989   0.846  0.3977
ParentChrom10 -1.75621    1.05543  -1.664  0.0961 .
ParentChrom11  0.63769    0.36566   1.744  0.0812 .
ParentChrom12 -0.87349    0.66495  -1.314  0.1890
ParentChrom13 -16.13169   961.32771 -0.017  0.9866
ParentChrom14 -0.48740    0.78800  -0.619  0.5362
ParentChrom15 -16.46909  1086.36946 -0.015  0.9879
ParentChrom16 -1.01779    0.78298  -1.300  0.1936
ParentChrom17 -1.70287    0.77891  -2.186  0.0288 *
ParentChrom18 -16.19108  1403.16051 -0.012  0.9908
ParentChrom19 -1.29083    0.66334  -1.946  0.0517 .
ParentChrom20 -0.75491    0.55328  -1.364  0.1724
ParentChrom21 -0.46597    1.06659  -0.437  0.6622
ParentChrom22 -16.22860  1824.77476 -0.009  0.9929
ParentChrom23 -0.62226    0.60394  -1.030  0.3029
ParentChrom3  -0.08735    0.52574  -0.166  0.8680
ParentChrom4  -0.29824    0.52664  -0.566  0.5712
ParentChrom5  -1.69828    1.05451  -1.610  0.1073
ParentChrom6   0.25217    0.48582   0.519  0.6037
ParentChrom7  -0.30828    0.45486  -0.678  0.4979
ParentChrom8 -16.11103   878.66590 -0.018  0.9854
ParentChrom9  -0.15793    0.52531  -0.301  0.7637
ParentChromX  -1.72131    0.78063  -2.205  0.0275 *
ParentChromY  -2.17207    1.05861  -2.052  0.0402 *
TypePSSD:TranscribedTRANS_YES  1.28185    0.59316   2.161  0.0307 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sex chromosomes are significant. Type is significant. Transcription has different influence on different types (But its marginal influence is not observable).

Number of Fisher Scoring iterations: 19

Worm - Fitted Models

- D=25K, no very significant contributors (independent variables).
- D=50K, Type starts to get significant.
- Consider all co-residence on the same chromosomes, nothing is significant except Chrom V.

Fly - Fitted Models

- $D=25K-100K$, Type is always and the only one significant.
- Consider all co-residence on the same chromosomes, nothing is very significant. (Type and Chrom3 a little sig.)

Exchange between sex chromosomes
and autochromosomes

Human: Exchange among X, Y and Autochromosomes (Hyp6)

As seen before:

		Parent (source)		
		Autochromosomes	X	Y
Pseudogene (destination)	Autochromosomes	7729	306	4
	X	549	81	1
	Y	105	33	87

- 3by3: Pearson's Chi-squared test & Fisher's Exact Test: p-value < 2.2e-16.
- 2by2 (X, Y only): Fisher's Exact Test: p-value < 2.2e-16.
Sample estimat: odds ratio \approx 208.
- Y chromosome very seldom exports to other chromosomes.


Worm: Exchange between X and Autochromosomes

Pseudogene (destination)		Parent (source)	
	Autochromosomes	Autochromosomes	X
	X	522	7
		19	6

- Fisher's Exact Test: p-value = 6.293e-06.
Sample estimat: odds ratio \approx 23.

Fly: Exchange among X, Y and Autochromosomes

		Parent (source)	
		Autochromosomes	X
Pseudogene (destination)	Autochromosomes	56	6
	X	2	29
	Y	4	0



- Fisher's Exact Test: p-value = 2.376e-16.
- Y chromosome very seldom exports to other chromosomes. (Y is not parent source.)